CHATPER 20

Solubility in Food, Pharmaceutical, and Cosmetic Industries

SIMÃO PEDRO PINHO[1] AND EUGÉNIA ALMEIDA MACEDO[2]

[1] *Laboratory of Separation and Reaction Engineering, Escola Superior de Tecnologia e de Gestão, Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5301-857 Bragança, Portugal*

[2] *Laboratory of Separation and Reaction Engineering, Departamento de Engenharia Química, Faculdade de Engenharia, Rua do Dr. Roberto Frias, 4200-465 Porto, Portugal*

20.1 Introduction

Solubility is well recognized as a fundamental physical property for the design of processes to separate, concentrate, and purify a targeted species. As will be discussed in the next section food, pharmaceuticals, and cosmetic industries frequently involve separation processes like precipitation, crystallization, liquid–liquid or supercritical fluid extraction (SFE). In each of these processes the choice of solvent plays an important role; for instance, it is estimated that 30% of the work of a thermodynamic group in a pharmaceutical company is directly related to the solvent selection.[1] As a result, we have decided to present this chapter in terms of the solubility of pharmaceuticals, amino acids (AA), proteins, or sugars in water, organic, and mixed solvents, liquid–liquid solubility; mostly related to aqueous two-phase systems (ATPS), or water–octanol partition coefficients, and solubility in supercritical fluids. The focus is, essentially, on the most recent developments concerning solubility correlation and prediction for substances of interest in those industries and processes. Experimental aspects, although of extreme relevance, are only highlighted for some specific cases where it is important to be aware of some particularities. In fact, several issues about the measurement of solubility were recently subject to an important edition.[2] Finally, a global overview is presented, some suggestions emphasized, and also some challenges for the near future are pointed out.

20.2 Industrial Importance

Recently, Agrawal and Noble[3] addressed some problems concerning separation needs for the 21st century. Many of them are related to the pharmaceutical, biomedical, and other biotech industries. In this context, researchers from Dow Chemical Company pointed out crystallization, ATPS, and other similar liquid–liquid extractions as processes of highest relevance.[4]

Excluding ethanol, antibiotics and AA are the major fermentation products with a market value around US$ 8 billion in 2004.[5] Their application in pharmaceutical or food industries are numerous, and AA are also used in the cosmetic industry. For instance, serine is employed for skin care cream or lotion, and some histidine derivatives act as free anti-radical agents in cosmetics.[6] After fermentation, several purification and separation techniques are applied to those highly complex broths. Crystallization is often used, for example, in glutamic acid or threonine production, for which solubility data is fundamental. Apart from key separation issues like extraction and crystallization, in pharmaceutical industries, solubility is also an essential property for the design of new drugs. Aqueous solubility gives valuable indications about the biological activity of a drug, and therefore, is most important in preformulation studies.[7] Water solubility, co-solvency and partition coefficients are topics under attention in many research and development groups at companies like Mitsubishi Chemical Corporation,[1] Merck and Zeneca Pharmaceuticals,[8] Hoffmann-La Roche,[9] GlaxoWellcome,[10] and Pfizer,[11] to name a few.

In the area of food processing, Agrawal and Noble[3] focused on solving such problems as the requirements of extremely high purity, and flavor and aroma capture. One of the technologies most studied in this area is the SFE with several patents and applications; some examples are the removal of cholesterol from food products,[12] de-alcoholization of beverages,[13] and concentration of flavor compounds.[14] However, even if the final sensory appreciation of flavor and aromas in food are much dependent on how the components are distributed over the different phases,[15] phase equilibria in food product design is still creating its basic foundations. Bruin,[15] and researchers at Unilever Research, applied a simple 2- or 3-suffix Margules equation[16] for the representation of the solid–liquid equilibrium of three polymorphic forms of fat crystals, sharing its success with the other well-known case, the solidification of chocolate.[17] A few final examples about research carried out for industrial needs are listed on Table 1.

20.3 Water Solubility

Water is omnipresent in many reaction and separation processes in biotechnology, and as discussed previously, solubility of biomolecules is a key equilibrium property in their production. Additionally, drug solubility in water gives general trends for rates of dissolution; poor solubility is usually

Table 1 *Some projects, involving solubility issues, carried out at different companies*

Problem Addressed	*Company*	*Ref.*
Effect of α-tocopherol on the solubilization of poor soluble drugs in simulated intestinal fluids	Dumex–Alpharma	Nielsen *et al.*[18]
Enrichment of Amaranth seed oil on high value lipids by SFE	Unilever	Westerman *et al.*[19]
Find an efficient excipient for rapamycin (immunosuppressor)	Schering–Plough HealthCare	Simamora *et al.*[20]
Increase the average crystal size of pharmaceuticals or agrochemicals by batch crystallization	Rhone–Poulenc	Lewiner *et al.*[21]
Influence of water content on triglycerides and their ability to be used as pharmaceutical excipients of steroids	Proctor and Gamble Novartis Pharmaceuticals	Land et al.[22]
Study of the solubility and partition coefficients of surfactants in several solvent systems to design initial extraction processes	Merck	Pollard *et al.*[23]
Study of ethanol as co-solvent in the crystallization of 1,3-dihydroxyacetone for application in the cosmetic industry	Ard–Soliance	Zhu *et al.*[24]

synonymous with a very low dissolution velocity.[25] As a result, an administered drug will mostly be excreted without the possibility of absorption from gastrointestinal tract into the cardiovascular system.[26] Besides the inherent complexities with experimental measurements, for this type of molecules, accuracy and reliability are specially difficult to achieve, and measurements are particularly time consuming. So, methods to predict water solubilities are an important research subject, with an extraordinary value for drug design. This task is, however, challenging because biomolecules are often very complex; they possess high molecular weight, with two or more functional groups, leading to a variety of complex molecular interactions, and are often present in different structures or isomers.[27]

In order to satisfy conditions for satisfactory water solubility and membrane permeability, drugs need to have the right balance between polarity and hydrophobicity. Empirically, if log S(S is the drug aqueous solubility in mol dm^{-3}) is in the range between -1 and -5, its adequacy is accepted.[26] Several methods to calculate log S for drugs have been proposed, but the correlations based on physicochemical properties like the octanol-water partition coefficient (P_{ow}) and the melting point are currently of little use. In fact, as will be briefly discussed in Section 5, to calculate P_{ow} several reliable methods are known, but for the melting point the opposite is true, and several reasons may hamper its experimental measurement. Another approach is based on the group-contribution concept, for which probably the most familiar is the AQUAFAC method. However, it also has the disadvantage of needing the melting point,[28] and even if several other methods that avoid this problem are available, generally, the number of groups is not enough to represent the wider variety

of drugs under development nowadays. Multiple linear regression (MLR) and neural networks (NN) are two other techniques applied for solubility predictions. They are both based on a set of different descriptors like molecular weight, solvent-accessible surface area, and many other topological and electronic indices. While NN allows the introduction of non-linearity for the descriptors terms in the solubility equation, which is an advantage to MLR, it is a black box type method that cannot provide insights for drug lead optimization except by trial and error.[26] In addition, over-training is a major issue for NN techniques, and its predictive capabilities are, most of the times, no better than that of MLR. Several different equations have been proposed and reviewed,[26,28,29] but many of them do not consider AA or sugars, and some molecular descriptors are not easy to understand physically. The linear-solvation energy relationship developed by Hoover and collaborators[30] is one of the most useful and comprehensive; equations have been derived for about 50 solvents and molecular descriptors for more than 3000 common organic and pharmaceuticals compounds have been calculated for solubility predictions. Recently, Sun[9] presented a more universal method, proposing atom types molecular descriptors to build predictive models for different properties, including log *S*.

These methods are all difficult to compare since they are based on different sets of experimental data. The usual strategy is to evaluate the predictive ability of the different methods to a test set, constituted by 21 different drugs and pesticides. Very good results on deviations for log *S* were obtained using the NN[28] and Sun[9] methods. As far as accuracy is concerned one cannot ask for much better results since it strongly depends on the uncertainty of the experimental measurements of log *S* which, for complex molecules, is around 0.6 log unit.[26]

Experimental uncertainty may be attributed to substance purity, different aspects related to the solid phase, pH and temperature control, and the method used. Givand and collaborators[31] developed studies on the influence of isomorphic impurities in the crystal purity of AA which is intimately linked to the relative solubility ratio and type of solvent. Other studies emphasize the importance of the solid-phases analysis[32,33] and the method chosen for the measurements. In fact, as shown in Figure 1, the equilibrium solubilities obtained by the cooling and the isothermal experimental methods present quite different results for the L-isoleucine+L-valine+water system at 298 K. Analyzing the solid phase, and applying mass balances, the authors concluded that the cooling method gives more consistent results.

Several different models have been proposed to represent thermodynamic properties of aqueous solutions of AA, namely, solubility and activity coefficients, with or without a presence of an electrolyte.[34] Although some progress has been achieved, the complexities that arise from the zwitterionic nature of AA in aqueous solution, make it a difficult task, and generally, it is not possible to calculate accurately the solubilities using activity coefficient data only. The fact that AA are the building blocks of more complex molecules such as antibiotics, peptides, or proteins, makes the understanding of the effect of

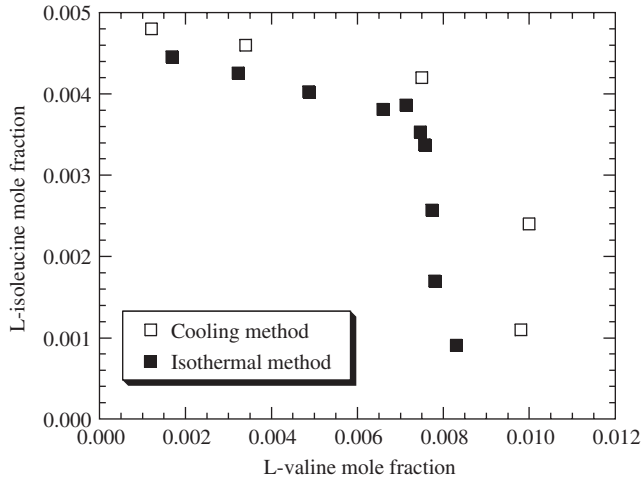

Figure 1 *Solubility in the L-isoleucine+L-valine+water system at 298 K: comparison of the cooling and the isothermal experimental methods.*[32]

electrolytes on the properties of aqueous AA solutions very important and attractive, as it may give insights into processes such as salt-induced precipitation of proteins. Figure 2 shows the effect of KCl on the glycine solubility in aqueous solution at 298.15 K.

The experimental results presented by Khoshkbarchi and Vera[35] and Ferreira *et al.*[34] are considerably different, which, once more, stresses the need of careful experimental planning. For that particular system, Khoshkbarchi and Vera[35] applied an equation based on the perturbation theory to correlate their activity coefficient data, but concluded they had to use an empirical correction to explain the solubility behavior, based on the observed effect of KCl on the crystallographic form of the AA. Alternatively, Ferreira *et al.*[34] correlated the same activity coefficient data, but with a modified form of the Pitzer–Simonson–Clegg equations,[36] and predicted the solubility assuming unchanged solid phase. The prediction curve is also included in Figure 2, suggesting a higher adequacy of their measured solubility data. Hamelink *et al.*,[37] in their studies about the effect of NaCl on the activity coefficients of antibiotics could not find a difference in the crystallographic structure of the solid phase formed by precipitation from electrolyte antibiotic solutions to explain the solubility behavior.

These studies are all important for a proper understanding of complex systems involving biomolecules, and might be useful for the investigation on protein solubility and crystallization. These questions are correlated, and rather complex, since protein crystallization/solubility depends on many factors such as pH, ionic strength, salt or protein type, temperature, surface hydrophobicity, and charge distribution, *etc.*, but extremely useful to identify, rationally, the

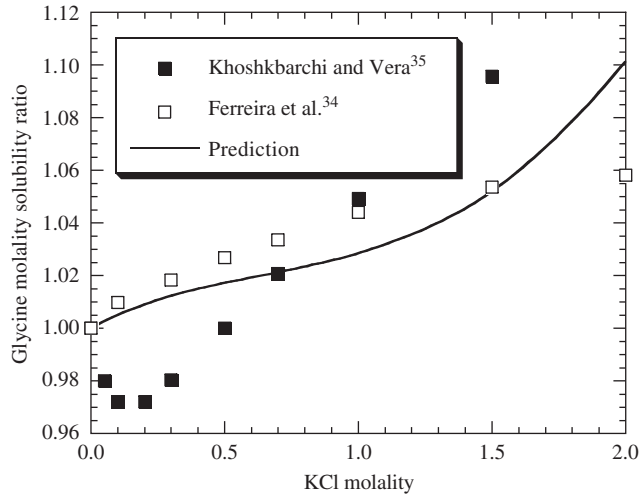

Figure 2 *Comparison of glycine solubility in water/KCl solutions at 298.15 K. The line represents prediction[34] using activity coefficient data only.*

optimal conditions for protein crystallization, reducing considerably the cost of a trial and error process. One interesting new effective predictive tool for protein crystallization is the introduction of the ''crystallization slot'' concept, which associates protein crystallization with the osmotic second virial coefficient (SVC–B_{22}). It can be briefly summarized in the following conservative way; while protein crystallization is very difficult for positive SVC values, it is favorable for negative values up to -10^{-3} mol ml g^{-2}, but do not guarantee successful crystal growth.[38] Although SVC is a thermodynamic property of dilute protein solutions, Guo *et al.*[39] have shown experimentally that it is also correlated with protein solubility. Figure 3(a) shows the surprising results when plotting these two variables for aqueous solutions of lysozyme obtained at different pH, temperature, salt type and concentration.

The link between those experimental observations and theory has been carried out by Haas *et al.*,[43] who used two different protein interaction potentials, and Rupert *et al.*,[44] who derived a two-parameter correlation based on classical thermodynamics, to represent the relation between solubility and SVC changing composition, temperature, or pH. Experimental determination of SVC by different methods like static or dynamic light scattering,[45,46] self-interaction,[38,41] or size-exclusion,[38] chromatography, can give, however, different values for the same protein under the same conditions. Figure 3(b) gives the SVC for lysozyme at different NaCl concentrations presenting considerable differences. The subject is delicate, since anisotropy effects are much relevant,[47] for instance, the substitution of a single AA in a protein may introduce big changes in the SVC values measured. Therefore, in this active research area it will be fundamental to have the development of more reliable methods, and the

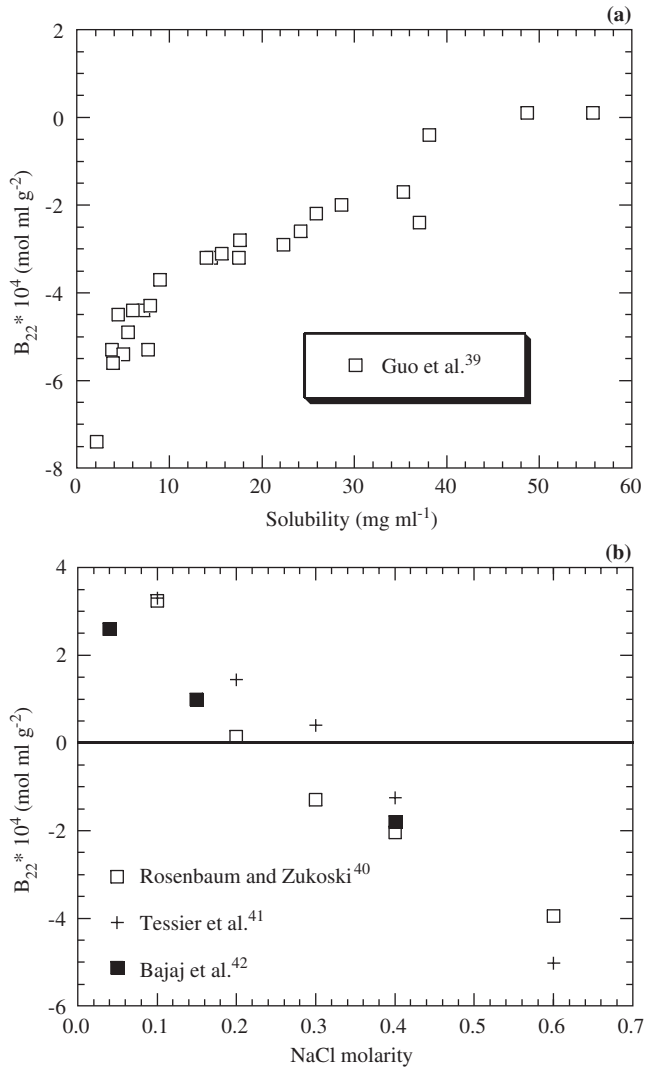

Figure 3 *(a) Experimental correlation between SVC and solubility for aqueous solutions of lysozyme; (b) Comparison of experimental results for SVC in aqueous lysozyme solutions at different NaCl molalities.*

extension of the conclusions for different proteins. Two different modeling approaches worth mention are the use of the UNIQUAC[48] equation to model protein solubility, and the neural network technology for protein crystallization, recently reviewed by DeLucas *et al.*[49] The lecture given by Prausnitz[50] on molecular thermodynamics for proteins in aqueous solution is highly recommended.

20.4 Organic and Mixed Solvent Solubility

Either for extraction, crystallization, or drug formulation purposes the study of co-solvency is common in pharmaceutical companies. Similarly, for water solubility, several methods have been proposed to calculate, and predict, the solubility of biomolecules in organic or mixed solvent systems. One attractive approach is the so-called log-linear model;[11] it presents two specific co-solvent parameters, and as far as the substance P_{ow} and water solubility are known, the solubility of a drug can, in principle, be estimated for an aqueous mixed solvent system. It has, however, a major deficiency as it cannot predict, or even correlate, solubility in systems like water/ethanol with caffeine, that present maxima[51] over the whole solvent composition range. The application of group-contribution methods is an alternative, but many group interactions relevant for pharmaceutical compounds are missing. Thus, the MOSCED (Modified Separation of Cohesive Energy Density) developed by Lazzaroni *et al.*[52] is a good alternative since group interaction parameters are not needed. Using a large number of data for activity coefficients at infinite dilution (γ^{∞}), 5 parameters were correlated for each of the 133 solvents studied. The MOSCED parameters for a given drug can easily be obtained if a few binary solubility data (the authors suggest 5–8 data points in chemically diverse solvent set) are available. After those are used to calculate the γ^{∞}'s, and from their values the Wilson or UNIQUAC[16] interaction parameters can be obtained, making possible the calculation of the solubility in all the mixed solvent composition range. A major drawback, as explained before, is that the melting properties must be known, and most probably, for many solutes, the data used to obtain the MOSCED parameters are too far from infinite dilution conditions. Nevertheless, for 26 solutes studied, an average absolute deviation (AAD) of 24.9% was found in the correlation of 700 solubility data points. Another method, perhaps one of the most used in the pharmaceutical industry, is the regular solution theory,[16] where the solubility is a function of the solvent solubility parameter. Often, a maximum in solubility is found, which corresponds both to the ideal solubility, and to the equality between solvent and solute solubility parameters. Again, solute-melting properties must be available, and even if for some solutes, like morphine in different solvents, the prediction is of high quality, for an hetero-atomic compound, the inadequacy of the method can be extremely pronounced.[1] In fact, in some very good solvents, the solute solubility can exceed significantly the ideal solubility, which is totally impossible to predict with the model.

AQ1

Avoiding some of the disadvantages pointed out in the previous paragraph, Abildskov and O'Connell[27] developed an ingenious reference solvent methodology. It involves the selection of a solvent, the "optimal solvent", which allows the calculation of the solute solubility in another solvent so long as the solubility in the optimal solvent and a predictive activity coefficient model, are available. In practice, the optimal solvent is found by a trial and error procedure, minimizing the difference (for a set of solvents) between the experimental solubility in a given solvent, and that calculated for the same solvent

using the reference solvent approach. The UNIFAC[16] method was selected to calculate the activity coefficients, and for cases where the interaction parameters are unknown, a sensitivity analysis in terms of the more relevant parameters is suggested, reducing considerably the experimental measurements needed. The results are really promising except, perhaps, when the solubility is very high. Extensions for mixed solvent systems,[53] and the inclusion of the temperature influence on the solubility temperature dependency[54] were recently proposed.

For the special case of amino acids, Orella and Kirwan[55] first suggested the use of the excess solubility approach to correlate the solubility of several amino acids in water/propanol and water/isopropanol mixtures with the Wilson model, obtaining an average relative deviation (ARD) of about 15.3%. Following on, Gude *et al.*,[56,57] used the same approach, but combining the Flory–Huggins (FH) theory with a Margules residual expression. Their method is very simple and attractive since the authors claim the use of a unique specific Margules parameter for each amino acid in all aqueous alkanol solutions, which allows a straightforward prediction of amino acid solubilities in alkanol/water solvents systems. However, applying their method to the description of the solubility of amino acids in water/methanol solvents, which are usually the easiest to correlate, the ARD found was 27.7%. To the best of our knowledge, the work by Ferreira *et al.*,[58] is the more comprehensive in this subject. Within the framework of the excess solubility approach, the NRTL model was applied for the correlation of the solubility of a large number of amino acids in several alkanol/water solvents. The temperature effect was included for some specific amino acids, and some predictions were made. The ARDs were 8.4% for correlation and 15% for predictions. Figure 4(a) compares the results achieved by Gude *et al.*[56] using the FH+Margules approach, with the NRTL results obtained by Ferreira *et al.*,[58] for the ratio between the solubility of the AA in the mixed solvent to that in pure water (relative solubility). A better agreement was found with the NRTL model for the solubility of the AA in aqueous 1-butanol solutions. Figure 4(b) shows the very good results for the prediction of glycine solubility in aqueous ethanol solutions at two different temperatures outside the temperature range used in the correlation.

Regarding carbohydrates, the increasing interest for food technology applications caused a great demand for predictive methods for both aqueous and mixed solvent solutions. In the last decade two kinds of approaches were proposed in the literature: molecular models and group-contribution methods.[59] Two modified UNIQUAC equations are available: the model presented by Peres and Macedo,[60] that uses fewer parameters for each sugar–water pair and adopts the symmetric convention, and allows a straightforward extension to mixed solvent systems. This is not possible with the other model suggested by Catté *et al.*[61] These authors chose the unsymmetric convention for the activity coefficients calculations. The major trend in recent modeling research is, however, based on the group-contribution methodology.

Different UNIFAC-based models are available for the prediction of solubilities in sugar solutions.[59] Some of the UNIFAC parameters have even been

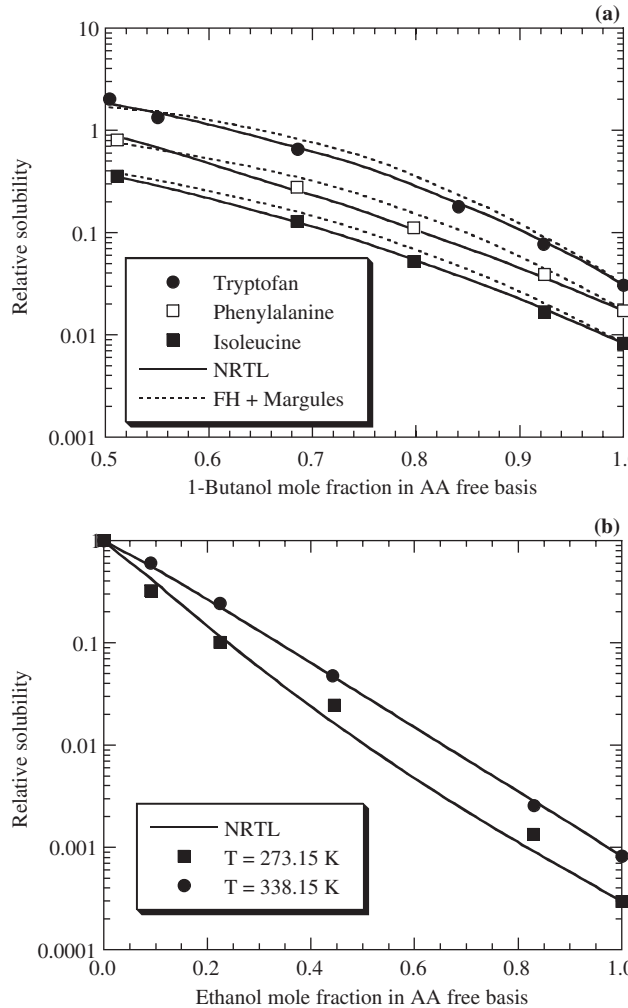

Figure 4 *(a) Relative solubilities of amino acids in water/1-butanol solutions at 298.15 K: comparison between the NRTL[58] and FH+Margules[56] models; (b) NRTL[58] predictions for the relative solubilities of glycine in water/ethanol solutions.*

predicted theoretically with methods of molecular mechanics.[62] The drawback of these models is the lack of accuracy at very high sugar concentrations (> 90%wt), as has recently been pointed out.[63] The reason for this lies in the fact that the majority of the data available does not cover this range of composition. To improve predictions of solubility in sugar solutions at these ranges of composition, new data were measured and a four-suffix Margules equation with temperature dependent parameters was presented in the literature,[63] as well as a new physical–chemical model.[64] This model takes account for the

hydration equilibrium of carbohydrates with the formation of carbohydrate *n*-water molecules and uses a UNIFAC model to describe the physical interactions. Although these very recent studies try to correct deficiencies found in other UNIFAC-based models, it is recognized[63] that the A-UNIFAC method developed by Ferreira *et al.*[65] is the tool with stronger theoretical foundations, allowing for a better capacity in predictive calculations. It incorporates a specific association term, which considers hydrogen bonding for sugar, water and other solvents molecules.

This section cannot be concluded without a brief mention of the innovative features of the NRTL-SAC model proposed by Chen and Song.[66] In this model the liquid non-idealities are described in terms of three types of conceptual segments of the molecules; hydrophobic, polar, and hydrophilic. Using reference substances for each type of segment, (hexane, water, and acetonitrile, respectively) an extensive binary VLE and LLE database, focused on the 62 solvents most used in the pharmaceutical industry, was used to estimate the number of conceptual segments required in each solvent. Following on, with a few selected solubility data values of the target solute, its number of conceptual segments can be calculated readily, and the solubility prediction in other solvents and mixed solvents is straightforward. As it requires some well-chosen data, NRTL-SAC is, like MOSCED and the reference solvent method, a hybrid-data estimation method that should be encouraged.[67] Its ability to model complex pharmaceuticals organic electrolytes has been already demonstrated,[68] and the potentialities to describe solubility of other types of solutes seems immense.

20.5 Liquid–Liquid Solubility

In the previous sections the importance of P_{ow} as a fundamental parameter for the estimation of solubilities in a variety of solvents has been stressed. Thermodynamics and extra-thermodynamics aspects of partitioning as well as its experimental and calculating methods were recently carefully reviewed by Sangster.[69] Owing to the uncertainty in the experimental P_{ow} values, Sangster also presents a list of recommended values for about 500 organic compounds. Thus, only the review by Derawi *et al.*[70] on group-contribution methods is briefly focused. Five different UNIFAC-based methods were compared, and the WATER UNIFAC,[71] and UNIFAC LLE[72] were recommended. These models, however, present a small number of interaction parameters available, and this inhibits their application for some functional groups like amines. For highly hydrophobic compounds, all the UNIFAC models underestimate P_{ow}, and generally, for AA, their derivatives, and sugars, P_{ow} is overestimated. The authors believe that for multifunctional compounds the group-contribution concept has limited capacity for further developments, and also that the atom/fragment correlation (AFC) method[73] showed superior performance in all cases studied. This method, similarly to the one proposed by Sun[9] (Section 3), allows the calculation of both P_{ow} and solubility by building a substance from atom

descriptors. In the AFC method MLR was applied to derive fragment coefficients and correction factors using 2473 P_{ow} in the training set, and around 10,600 for the validation of the method. The results seem really remarkable as it is possible to take into account steric interactions, hydrogen bondings, and even for zwitterionic species like ampicillin, amoxycillin, or peptides, values of P_{ow} can be estimated. A free online interactive demonstration to calculate P_{ow} is available at http://www.syrres.com/esc/kowdemo.htm.

Despite the increase and progresses achieved in the research work on ATPS, so far the studies are rather scattered, making the knowledge of the mechanisms of solute partitioning, limited. This is probably one of the main reasons for the reluctance in its commercial exploitation.[74] Traditionally, protein partitioning has been studied in polyethyleneglycol (PEG)/dextran or PEG/(phosphate or sulfate) salt, and the factors to consider, beyond those mentioned earlier for protein crystallization, must now include some characteristics of the polymer(s). The implementation of general rules to choose the best ATPS and the best operating conditions for a given separation, will make practical applications simpler. However, making those rules accessible depends much on how these different factors are understood. Recently, some interesting attempts have been made: Lin *et al.*[75] studied the influence of polymer concentration and molecular weight; Andrews *et al.*[76] focused on the protein charge and surface hydrophobicity, which was also done by Tubio et al.[77] However, no general trend was found. Even if a relationship between the hydrophobic character of the partitioned substance and its partitioning coefficient was found the general picture is, when studying polymer molecular weight effects the conclusions are limited to certain proteins, and studying the protein surface hydrophobicity effects, the results are restricted to certain values of the polymer molecular weight.

Though much more experimental work is needed, the application of molecular thermodynamics to this kind of problems must have the highest priority. In the recent past, several different approaches have been proposed concerning protein, peptides, and AA partition in ATPS. This was recently reviewed briefly by Jiang and Prausnitz,[78] who also derived a model that takes into account, successfully, the different partitioning behavior of native and denatured proteins. One of the most recent studies on protein partitioning, and perhaps the most comprehensive, is due to Madeira *et al.*[79] Their modified Wilson model, based on the lattice theory and the two-fluid theory, was successively applied to the representation of electrolyte solutions, water activity in aqueous polymer solutions, and polymer/polymer or polymer/salt ATPS. A Debye-Hückel term was included to take into consideration the long-range nature of the electrostatic forces in solution, and the authors end up with a model where only the interactions involving proteins are needed to calculate protein partitioning. To simplify, Madeira *et al.*[79] fixed those at zero, and calculated the partition of four different proteins in Na_2SO_4/PEG6000 and K_2HPO_4/PEG6000 by adjusting the protein net charge. Globally, the results may be considered very reasonable even if in some cases large discrepancies were found between the experimental and the calculated net charge. That is not the case for the

partitioning behavior of lysozyme in K_2HPO_4/PEG6000 aqueous system at 298.15 K shown in Figure 5. Here the published experimental value for the net charge is two, and it produces much higher deviations on the calculated partition coefficient than that obtained using the fitted value of four for the net charge. The complexity of the problem and the lack of data remain as the major reasons for the development of more efficient predictive tools for protein partitioning on ATPS. However, some useful insights from protein crystallization must also be considered, and as it is expected that ATPS will be extend into food and cosmetic industries,[74] these problems will continue to draw attention in the near future.

20.6 Solubility in Supercritical Fluids

Contrary to the work on the ATPS, most of the studies on SFE started in the area of food technology. In fact, around 125 industrial scale SFE units are in operation,[80] and some of those applications were reviewed by Knox.[81] The studies are now spreading into the area of drug processing that is currently a very popular research area, namely, on the purification, crystallization, or micronization of pharmaceuticals. In this context, a review on techniques such as rapid expansion supercritical solution, or gas anti-solvent system is given

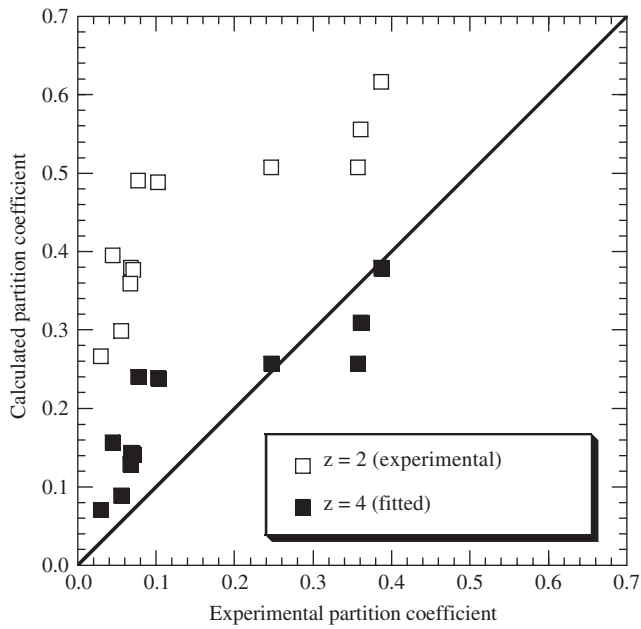

Figure 5 *Influence of the protein net charge (z) on the lysozyme partition coefficient calculation, by a modified Wilson model,[79] in K_2HPO_4/PEG6000 aqueous system at 298.15 K.*

eslsewhere.[80,82]

The large majority of the studies concern the solubility of a solute in carbon dioxide. Some supercritical measurements on other systems such as the solubility of solid mixtures and the solubility in solvents other than carbon dioxide (and co-solvency) are also available.[83–85] To correlate solubilities, empirical equations have usually been applied, and the Chrastil equation[86] is one of the most often used. Several other empirical equations have been proposed through the years, and some were recently compared for solute solubility in carbon dioxide by Jouyban *et al.*[87] Avoiding the difficulties of calculating some physicochemical properties, the authors only compared equations for which the independent variables are readily accessible, *e.g.*, temperature, pressure, and pure solvent density promoting, in this way, their usefulness. A six parameter equation, developed by the authors, showed the best performance with an AAD of 21.4% for the 106 systems compiled in their database. This is comparable to the experimental uncertainty.

Several different equations of state (EoS) have also been applied, but a major difficulty arises from the need to know solute critical properties, vapor pressure, and density. Unfortunately, for many substances those are impossible to measure because the solute decomposes, and estimating methods must be applied for their evaluation. However, different methods give, naturally, different values for those properties, which may have a lot of influence in the correlation abilities of the EoS and, what is worse, can produce poor and sometimes impossible results.[88] Furthermore, relatively small variations in the properties can cause large differences in the predicted solubilities,[89] and so, a lot of caution must be taken in the choice of methods used. Another issue that must be considered carefully when modeling solubilities is the stability of the results. In a very enlighten work, Xu *et al.*[90] developed a strategy, and gave some good examples, about the need of considering the iso-fugacity condition simultaneously with a check on global thermodynamic phase stability by applying tests such as tangent plane analysis and global minimization methodologies. Cubic EoS like Soave–Redlich–Kwong and Peng–Robinson are, surely, the most used, but for rigorous calculations of solubilities in these type of systems much more work is imperative. Nevertheless, even if accurate general conclusions are not possible at this time, cubic EoS that uses free energy models in its parameters and non-quadratic mixing rules, with interaction parameters in the volume constants, give the best results.[91] Finally, taking into consideration the results already achieved with the group-contribution associating EoS,[92] the research on the potential of association fluid theories is also highly recommended.

20.7 Conclusions

A global overview about current solubility issues for food, pharmaceutical, and cosmetic industries has been given. Great progress has been achieved for solute solubility in water and organic solvents as well as for water–octanol partition

coefficients, but the potentialities of some very recent models and methods, *e.g.*, reference solvent methodology or NRTL-SAC model, should be extensively explored. Nevertheless, constant evolution in those industries will stress the need for new measurements and advances for innovative experimental techniques. The development of a solubility database and a measurement strategy, perhaps, as suggested by Kolář *et al.*,[1] is highly recommended, but applications of relatively novel compounds like ionic liquids or cyclodextrins should also be taken into account for the development of new processes. A very interesting progress on protein crystallization has also been achieved with the "crystallization slot" concept that should be applied to several different systems. Understanding the behavior of simple molecules like AA and small peptides in aqueous electrolyte solutions can also be useful for further developments. As far as ATPS are concerned some interesting studies have briefly been discussed. However, the development of alternative ATPS as well as much more informative models capable of explaining mechanisms under protein partitioning is fundamental to make the technique attractive to industries. That is also an issue for simulation SFE processes, but EoS for associating fluids might be a very useful tool. In fact, some particular difficulties pointed out for solubility modeling in supercritical fluids make it a hard task, and an analysis for methods to estimate solute properties must be taken into account. Finally, it is consensual that, generally, understanding solubility phenomena may benefit very much from molecular simulation data.

20.8 Uncited References

40; 42

References

1. P. Kolář, J.-W. Shen, A. Tsuboi and T. Ishikawa, *Fluid Phase Equilib.*, 2002, **194–197**, 771.
2. G.T. Hefter and R.P.T Tomkins, (eds), *The Experimental Determination of Solubilities*, Wiley, Chichester, 2003.
3. R.D. Noble and R. Agrawal, *Ind. Eng. Chem. Res.*, 2005, **44**, 2887.
4. S. Gupta and J.D. Olson, *Ind. Eng. Chem. Res.*, 2003, **42**, 6359.
5. W. Leuchtenberger, K. Huthmacher and K. Drauz, *Appl. Microbiol. Biotechnol.*, 2005, **69**, 1.
6. K. Araki and T. Ozeki, Amino acids, *Kirk-Othmer Encyclopedia of Chemical Technology*, Wiley, 2003.
7. A. Blasko, A. Leahy-Dios, W.O. Nelson, S.A. Austin, R.B. Killion, G.C. Visor and I.J. Massey, *Monatsh. Chem.*, 2001, **132**, 789.
8. B. C. Hancock, P. York and R. C. Rowe, *Int. J. Pharm.*, 1997, **148**, 1.
9. H. Sun, *J. Chem. Inf. Comput. Sci.*, 2004, **44**, 748.
10. M.H. Abraham, J.A. Platts, A. Hersey, A.J. Leo and R.W. Taft, *J. Pharm. Sci.*, 1999, **88**, 670.

AQ2

11. J.W. Millard, F.A. Alvarez-Núñez and S.H. Yalkowsky, *Int. J. Pharm.*, 2002, **245**, 153.
12. R. Hartono, G.A. Mansoori and A. Suwono, *Chem. Eng. Sci.*, 2001, **56**, 6949;and references therein.
13. F. J. Señoráns, A. Ruiz-Rodríguez, E. Ibáñez, J. Tabera and G. Reglero, *J.Supercrit. Fluids*, 2003, **26**, 129.
14. Z. Shen, V. Mishra, B. Imison, M. Palmer and R. Fairclough, *J. Agric. Food Chem.*, 2002, **50**, 154.
15. S. Bruin, *Fluid Phase Equilib.*, 1999, **158–160**, 657.
16. J.M. Prausnitz, R.N. Lichtenthaler and E.G. Azevedo, *Molecular Thermodynamics of Fluid-Phase Equilibriua*, 3rd edn, Prentice-Hall, Upper Saddle River, NJ, 1999.
17. T. Morwood, *CAPE, Eureka Project 2311*, 2001.
18. P.B. Nielsen, A. Müllertz, T. Norling and H.G. Kristensen, *Int. J. Pharm.*, 2001, **222**, 217.
19. D. Westerman, R.C.D. Santos, J.A. Bosley, J.S. Rogers and B. Al-Duri, *J. Supercrit. Fluids*, 2006, **37**, 38–52.
20. P. Simamora, J.M. Alvarez and S.H. Yalkowsky, *Int. J. Pharm.*, 2001, **213**, 25.
21. F. Lewiner, G. Févotte, J.P. Klein and F. Puel, *Ind. Eng. Chem. Res.*, 2002, **41**, 1321.
22. L. M. Land, P. Li and P.M. Bummer, *Pharm. Res.*, 2005, **22**, 784.
23. J. M. Pollard, A. J. Shi and K. E. Göklen, *J. Chem. Eng. Data*, 2006, **51**, 230.
24. Y. Zhu, D. Youssef, C. Porte, A. Rannou, M.P. Delplancke-Ogletree and B.L.M. Lung-Somarriba, *J. Cryst. Growth*, 2003, **257**, 370.
25. R.H. Muller and C.M. Keck, *J. Biotechnol.*, 2004, **113**, 151.
26. W.L. Jorgensen and E.M. Duffy, *Adv. Drug Deliv. Rev.*, 2002, **54**, 355.
27. J. Abildskov and J.P. O'Connell, *Ind. Eng. Chem. Res.*, 2003, **42**, 5622.
28. J. Huuskonen, *Comb. Chem. High Throughput Screening*, 2001, **4**, 311.
29. J.S. Delaney, *Drug Discov. Today*, 2005, **10**, 289.
30. K.R. Hoover, W.E. Acree, Jr. and M.H. Abraham, *J. Solution Chem.*, 2005, **34**, 1121.
31. J.C. Givand, A.S. Teja and R.W. Rousseau, *AIChE J.*, 2001, **47**, 2705.
32. I. Kurosawa, A.S. Teja and R.W. Rousseau, *Ind. Eng. Chem. Res.*, 2005, **44**, 3284.
33. I. Kurosawa, A.S. Teja and R.W. Rousseau, *Fluid Phase Equilib.*, 2004, **224**, 245.
34. L.A. Ferreira, E.A. Macedo, E.A. and S.P. Pinho, *Ind. Eng. Chem. Res.*, 2005, **44**, 8892; and references therein.
35. M.K. Khoshkbarchi and J.H. Vera, *Ind. Eng. Chem. Res.*, 1997, **36**, 2445.
36. S.L. Clegg and K.S. Pitzer, *J. Chem. Phys.*, 1992, **96**, 3513.
37. J.M. Hamelink, E.S.J. Rudolph, L.A.M. van der Wielen and J.H. Vera, *Biophys. Chemist.*, 2002, **95**, 97.
38. T. Ahamed, M. Ottens, G.W.K. van Dedem and L.A.M. van der Wielen, *J. Chromatogr. A*, 2005, **1089**, 111.

39. B. Guo, S. Kao, H. McDonald, A. Asanov, L.L. Combs and W.W. Wilson, *J. Cryst. Growth*, 1999, **196**, 424; and references therein.
40. D. F. Rosenbaum and C. F. Zukoski, *J. Cryst. Growth*, 1996, **169**, 752.
41. P.M. Tessier, A.M. Lenhoff and S.I. Sandler, *Biophys. J.*, 2002, **82**, 1620.
42. H. Bajaj, V.K. Sharma and D.S. Kalonia, *Biophys. J.*, 2004, **87**, 4048.
43. C. Haas, J. Drenth and W.W. Wilson, *J. Phys. Chem. B*, 1999, **103**, 2808.
44. S. Ruppert, S.I. Sandler and A.M. Lenhoff, *Biotechnol. Prog.*, 2001, **17**, 182.
45. R.A. Curtis, J.M. Prausnitz and H.W. Blanch, *Biotechnol. Bioeng.*, 1998, **57**, 11.
46. W. Liu, T. Cellmer, D. Keerl, J.M. Prausnitz and H.W. Blanch, *Biotechnol. Bioeng.*, 2005, **90**, 482.
47. R. A Curtis and L. Lue, *Chem. Eng. Sci.*, 2006, **61**, 907.
48. S. M. Agena, M. L. Pusey and I. D. L. Bogle, *Biotechnol. Bioeng.*, 1999, **64**, 144.
49. L.J. DeLucas, D. Hamrick, L. Cosenza, L. Nagy, D. McCombs, T. Bray, A. Chait, B. Stoops, A. Belgovskiy, W.W. Wilson, M. Parham and N. Chernov, *Progr. Biophys. Mol. Biol.*, 2005, **88**, 285.
50. J.M. Prausnitz, *Pure Appl. Chem.*, 2003, **75**, 859.
51. P. Bustamante, J. Navarro, S. Romero and B. Escalera, *J. Pharm. Sci.*, 2001, **91**, 874.
52. M.J. Lazzaroni, D. Bush, C.A. Eckert, T.C. Frank, S. Gupta and J.D. Olson, *Ind. Eng. Chem. Res.*, 2005, **44**, 4075.
53. J. Abildskov and J.P. O'Connell, *Mol. Simul.*, 2004, **30**, 367.
54. J. Abildskov and J.P. O'Connell, *Fluid Phase Equilib.*, 2005, **228**, 395.
55. C.J. Orella and D.J. Kirwan, *Ind. Eng. Chem. Res.*, 1991, **30**, 1040.
56. M.T. Gude, H.H.J. Meuwissen, L.A.M. van der Wielen and K. Ch. A.M. Luyben, *Ind. Eng. Chem. Res.*, 1996, **35**, 4700.
57. M.T. Gude, L.A.M. van der Wielen and K. Ch. A.M. Luyben, *Fluid Phase Equilib.*, 1996, **116**, 110.
58. L.A. Ferreira, E.A. Macedo and S.P. Pinho, *Chem. Eng. Sci.*, 2004, **59**, 3117; and references therein.
59. E.A. Macedo, *Pure Appl. Chem.*, 2005, **77**, 559; and references therein.
60. A.M. Peres and E.A. Macedo, *Fluid Phase Equilib.*, 1996, **123**, 71.
61. M. Catté, C.G. Dussap, C. Achard and J.B. Gros, *Fluid Phase Equilib.*, 1994, **96**, 33.
62. S.Ó. Jónsdóttir and P. Rasmussen, *Fluid Phase Equilib.*, 1999, **158–160**, 411.
63. M. Starzak and M. Mathlouthi, *Food Chem.*, 2006, **96**, 346.
64. L.B. Gaïda, G.G. Dussap and J.B. Gros, *Food Chem.*, 2006, **96**, 387.
65. O. Ferreira, E.A. Brignole and E.A. Macedo, *Ind. Eng. Chem. Res.*, 2003, **42**, 6212.
66. C.-C. Chen and Y. Song, *Ind. Eng. Chem. Res.*, 2004, **43**, 8354.
67. P. M. Mathias, *Fluid Phase Equilib.*, 2005, **228–229**, 49.
68. C.-C. Chen and Y. Song, *Ind. Eng. Chem. Res.*, 2005, **44**, 8909–8921.

69. J. Sangster, *Octanol-Water Partition Coefficients: Fundamentals and Physical Chemistry*, Wiley, Chichester, 1997.
70. S.O. Derawi, G.M. Kontogeorgis and E.H. Stenby, *Ind. Eng. Chem. Res.*, 2001, **40**, 434.
71. F. Chen, J. Holten-Andersen and H. Tyle, *Chemosphere*, 1993, **26**, 1325.
72. T. Magnussen, P. Rasmussen and A. Fredenslund, *Ind. Eng. Chem. Process Des. Dev.*, 1981, **20**, 331.
73. W.M. Meylan and P.H. Howard, *Perspect. Drug Discov. Des.*, 2000, **19**, 67.
74. M. Rito-Palomares, *J. Chromatogr. B*, 2004, **807**, 3.
75. D.-Q. Lin, Y.-T. Wu, L.-H. Mei, Z.-Q. Zhu and S.-J. Yao, *Chem. Eng. Sci.*, 2003, **58**, 2963.
76. B.A. Andrews, A.S. Schmidt and J.A. Asenjo, *Biotechnol. Bioeng.*, 2005, **90**, 380.
77. G. Tubio, B. Nerli and G. Picó, *J. Chromatogr. B*, 2004, **799**, 293.
78. J. Jiang and J.M. Prausnitz, *J. Phys. Chem. B*, 2000, **104**, 7197; and references therein.
79. P.P. Madeira, X.Xu, J.A. Teixeira and E.A. Macedo, *Biochem. Eng. J.*, 2005, **24**, 147; and references therein.
80. M. Sihvonen, E. Järvenpää, V. Hietaniemi and R. Huopalahti, *Trends Food Sci. Technol.*, 1999, **10**, 217.
81. D. E. Knox, *Pure Appl. Chem.*, 2005, **77**, 513.
82. F. Dehghani and N. R. Foster, *Curr. Opin. Solid State Mater. Sci.*, 2003, **7**, 363.
83. M. Christov and R. Dohrn, *Fluid Phase Equilib.*, 2002, **202**, 153.
84. F.P. Lucien and N.R. Foster, *J. Supercrit. Fluids.*, 2000, **17**, 111.
85. W.H. Hauthal, *Chemosphere*, 2001, **43**, 123.
86. J. Chrastil, *J. Phys. Chem.*, 1982, **86**, 3016.
87. A. Jouyban, H.-K. Chan and N. R. Foster, *J. Supercrit. Fluids*, 2002, **24**, 19.
88. P. Coimbra, C.M.M. Duarte and H.C. Sousa, *Fluid Phase Equilib.*, 2006, **239**, 188.
89. G.I. Burgos-Solórzano, J.F. Brennecke and M.A. Stadtherr, *Fluid Phase Equilib.*, 2004, **220**, 57.
90. G. Xu, A.M. Scurto, M. Castier, J.F. Brennecke and M.A. Stadtherr, *Ind. Eng. Chem. Res.*, 2000, **39**, 1624.
91. J.O. Valderrama, *Ind. Eng. Chem. Res.*, 2003, **42**, 1603.
92. T. Fornari, A. Chafer, R.P. Stateva and G. Reglero, *Ind. Eng. Chem. Res.*, 2005, **44**, 8147.